J.King Saud Univ., Vol.8, *Computer & Information Sci.*, pp.21-52 (A.H 1416/1996)

An End-Case Analyzer of Arabic Sentences

A. D. Al-Sawadi and M. G. Khayat

**KFUPM, Dhahran, Saudi Arabia, **Dept. of Electrical & Computer Engineering*
King Abdul Aziz University, Jeddah, Saudi Arabia

(Received, 06 Sept. 1993; accepted for publication, 21 Sept. 1994)

Abstract. Natural Language Understanding (NLU) has been a growing area of research in computer science. Although morphology and syntax play an essential role in NLU, end-case analysis in some languages, like Arabic, is an important component in the correct interpretation of a sentence.

In this research, an end-case analyzer (ECA) for Arabic sentences has been designed, implemented, and integrated within a Natural Arabic Understanding System (NAUS). ECA consists of two main components: the end-case generator, and the invoker. The end-case generator determines the end-case analysis for the given sentence and its constituents according to Arabic end-case grammar rules, which have been encoded in ECA as IF-THEN Prolog rules and predicates. The invoker was implemented as calls to the end-case generator inserted within the syntactic analyzer of NAUS. ECA was implemented in Prolog on a personal computer with Arabic support.

The modular design of ECA permits its use within NAUS in addition to other applications such as sentence generation, speech analysis and synthesis, translation systems, teaching Arabic, and checking and correcting grammatical errors.

Introduction

The study of the Arabic sentence structure includes two main aspects: syntax and end-case. In Arabic, the determination of the syntactic identity and end-case

characteristics is called "irab". When parsing a sentence, end-cases play an essential role in determining the syntactic functions of the elements (words, phrases, clauses) of the sentence. Although morphological analysis determines morphological characteristics such as tense, affixes, gender, number, person and root, it can not determine the syntactic role or the end-case specially when the end-case is a diacritic, which is not explicitly specified in most Arabic text. Fig.1 shows two examples, the first of which contains two occurrences of a word that have the same exact morphological characteristics; yet their syntactic roles and end-case are both different. The second example shows a sentence where the end-cases are diacritics and the sentence has two different interpretations depending on the end-cases.The example shows the sentence with and without diacritics in both cases. End-case analysis is essential not only for understanding Arabic sentences and correct pronunciation of words but also for speech processing, translation systems, teaching Arabic, and checking and correcting grammatical errors.

In this research work, we present and end-case analyzer of Arabic sentences. We first present the linguistic principles of end-case in Arabic, define the problem, and survey the existing efforts in the field. Next, we present the proposed analyzer design, which is based on the linguistic principles of end-case, and the corresponding implementation, and demonstrate its use within a Natural Arabic Understanding System (NAUS). Finally, we conclude by evaluating the proposed analyzer, and considering different applications of the end-case analyzer.

Linguistic Principles of End-Case in Arabic

Words in Arabic are classified into three classes: nouns, verbs and particles [1 - 4]. The effect of end-case rules on a word can be an explicit or implicit mark. In the explicit case, the mark can be a diacritic, suffix, or deletion of the terminal letter of the word. In the implicit case, the mark is a diacritic which can not be uttered (thus implicit) due to phonological rules. In addition, end-case rules not only apply to words but also to subsentences (phrases, clauses and sentences), in which case the end-case mark is implicit. The end-case of a subsentence determines its syntactic function, thus contributing to the correct interpretation of the containing sentence.

Nouns can be in one of three irab cases: nominative, accusative, or genitive. A noun can be in one of these cases according to the end-case position that it occupies in a sentence. Tables 1a, 1b, and 1c show the possible end-case positions (syntactic function) and corresponding irab cases with examples. From the end-case point of view, however,

	End-case position		
Example 1	second part of possesive construct	ahdar ar-rajulu hagaa'ib [al-musaafereen] The man brought the luggage of the passengers	أحضر الرجل حقائب [المسافرين]
	object of verb	gabala ar-rajulu [al-musaafreen] The man met the passengers	قابل الرجل [المسافرين]
Example 2	subject of verb	ra'a [ar-rajulu] al-walada The man saw the boy	رأى [الرجلُ] الولدَ رأى [الرجل] الولد
	object of verb	ra'a [ar-rajula] al-waladu The boy saw the man	رأى [الرجلَ] الولدُ رأى [الرجل] الولد

Fig. 1. Examples

nouns in Arabic fall into one of two categories: variable and invariable [1 - 4]. Table 2 shows the two categories, the noun types in each category, and examples of each type.

A variable noun is a noun that experiences changes in its end-case mark when its irab case in the sentence changes [1 - 4]. The end-case mark for this type of noun differs from one noun type to another. That is, for some of these types the end-case mark is a diacritic, while for others it is a letter. Tables 3a, 3b, and 3c show a list, with examples, of these noun types and corresponding end-case marks.

An invariable noun in Arabic is a noun that does not experience any change in its end-case mark even when its end-case position in the sentence changes [1 - 4]. The end-case mark for this type of nouns is called, "al-bena'a" mark. This mark differs from one noun to another according to its terminal letter, but it is fiexed for the same noun whatever its irab case is. Since the end-case mark in this case is fixed, the noun's end-case is said to be implicit.

Verbs are classified into three classes: past, present, and imperative [2 - 4]. From another perspective, a verb in Arabic is either perfect or imperfect [2 - 4]. A perfect verb is one that must be followed by a subject of verb ("fael"), which may be explicit or implicit. An imperfect verb is one that is followed by a nominal sentence, which consists of a subject, ("mobtada"), and a predicate ("khabar"). When a nominal sentence is preceded by one of these verbs, the subject is renamed as the verb's subject, and the predicate is renamed as the verb's predicate. While the verb's subject retains its nominative case, the case of the verb's predicate becomes accusative. The third and fourth examples in Tables 1a and 1b respectively demonstrate the effect of the imperfect verb. Past and present tense Versbs are also divided into active, and passive voice. From the end-case point of view, however, verbs in Arabic fall into one of two categories: variable and invariable [1 - 4].

A variable verb is a verb that experiences changes in its end-case mark according to changes in its irab case in the sentence. This change occurs if the verb is preceded by a verbal particle. The present tense verb, that is not suffixed by the feminine "noon" or the energetic "noon" [5] is the only verb type that comprises this category of verbs. Variable verbs can be in one of three irab cases: indicative, subjunctive, and jussive [1], [2 - 4]. If a verb is preceded by one of the subjunctive particles, then the verb is said to be in the subjunctive case. If it is preceded by one of the jussive particles, then it is said to be in the jussive case. Otherwise, if none of these particles precedes a verb,then the verb is said to be in the indicative case. The end-case mark of a verb depends on its case,

Examples	الحالة الإعرابية: الرفع Irab Case: Nominative	Word End-Case Position الموقع الإعرابي للكلمة
الرجل كريم ar-rajulu kareemu	مبتدا subject	
الرجل كريم ar-rajulu kareemu	خبر predicate	
كان الرجل كريما kana ar-rajulu kareema	اسم كان وأخواتها subject of kana	
إن الرجل كريم inna ar-rajula kareemu	خبر إن وأخواتها predicate of inna	
لا أحد كريم la ahada kareemu	خبر لا النافية للجنس predicate of negation la	
ضرب الولد الكرة daraba al-waladu al-kurata	فاعل subject of active voice verb	
قُتِلَ الرجل gutela ar-rajulu	نائب فاعل subject of passive voice verb	
يا محمد ya mohammadu	منادى علم designated proper name vocative	
يا مغيث ya mugeethu	منادى نكرة مقصودة designated indefinite substantive vocative	
جاء الرجل الكريم ja'a ar-rajulu al-kareemu	تابع لمرفوع dependent of a nominative	

Table 1a. Nominative noun irab cases and the corresponding end-case positions

Examples	الحالة الإعرابية: النصب Irab Case: Accusative	Word End-Case Position الموقع الإعرابي للكلمة
ضربه ضربا شديدا darabahu darban shadeeda	مفعول مطلق absolute object	
مشيت و الرصيف mashaitu wa ar-raseefa	مفعول معه object of accompaniment	
خرج خالد صباحا kharaja khalidu sabaha	مفعول فيه adverbial object	
كان الطقس باردا kana at-tagsu barida	خبر كان object of kana	
إن الرجل كريم inna ar-rajula kareemu	اسم إن subject of inna	
لا أحد في المنزل la ahada fi al-manzili	اسم لا النافية للجنس predicate of negation la	
أخذ الولد الكرة akhatha al-waladu al-kurata	مفعول به object of verb	
ذهب الرجل ضاحكا thahaba ar-rajulu dahika	حال adverb of state	
أخذ ثلاثين قرشا akhatha thalaatheena qersha	تمييز distinction nouns	
يا طالب العلم اجتهد ya taliba al-elmi ejtahed	منادى نكرة غير مقصودة undesignated indefinite substantive vocative	
قام الرجل إكراما له gama ar-rajulu ikraman lahu	مفعول لأجله object of purpose	
أكل الحضور إلا الولد akala al-hodour illa al-walada	المستثنى exclusive adverbial exception noun	
ضرب الكرة الكبيرة daraba al-kurata al-kabeerata	تابع لمنصوب dependent of an accusative	

Table 1b. Accusative noun irab cases and the corresponding end-case positions

Examples	الحالة الإعرابية: الجر Irab Case: Genitive	
صاحب الدار كريم sahibu ad-dari kareemu	مضاف إليه second part of the possessive construct	الموقع الإعرابي للكلمة Word End-Case Position
خرج إلى الملعب kharaja ila al-mala'ab	مجرور بحرف جر object of preposition	
الرجل في البيت الصغير ar-rajulu fi al-baiti al-sagheer	تابع لمجرور dependent of a genitive	

Table 1c. Genitive noun irab cases and the corresponding end-case positions

Variable (معرب)	Invariable (مبني)
اسم العلم (أحمد ، علي ، ...) Proper names	اسم الموصول (الذي ، التي ، ...) Conjunctive (Relative) pronouns
الأسماء الخمسة (أبو ، أخو ، ...) The Five nouns	اسم الإشارة (هذا ، هذه ، ...) Demonstrative pronouns
الاسم المفرد (ولد ، رجل ، ...) Singular nouns	اسم شرط (إذا ، مهما ، ...) Conditional pronouns
مثنى (ولدان ، رجلان ، ...) Dual nouns	الضمائر (هو ، هي ، ...) Personal pronouns
جمع تكسير (رجال ، منازل ، ...) Broken plural nouns	اسم الفعل (هيهات ، أف ، ...) Verb exaggeration & sound nouns
جمع سالم (بنات ، مسافرون ، ...) Sound plural nouns	اسم استفهام (من ، أين ، ...) Interrogative nouns
اسم العدد غير المركب (عشرون ، ...) Non-compound numerals	اسم العدد المركب (تسعة عشر ، ...) Compound numerals

Table 2. Arabic noun types

الحالة الإعرابية وعلامة الإعراب Irab cases & End-Case Marks		نوع الاسم Noun Type	
Examples	الرفع Nominative		
كان الرجلُ هنا kana ar-rajulu huna	الضمة damma	الاسم المفرد بأنواعه singular	معرب بالحركات diacritic end-case mark
الأولادُ يلعبون al-awladu yala'aboon	الضمة damma	جمع التكسير broken plural	
البناتُ يلعبن al-banaatu yala'abna	الضمة damma	جمع المؤنث السالم sound feminine plural	
المسافرون مستعدون al-musaferoon musta'edoon	الواو waw	جمع المذكر السالم sound masculine plural	معرب بالحروف letter end-case mark
الولدان في المنزل al-waladaani fi al-manzili	الألف alef	المثنى dual	
أبو الولد كريم abu al-waladi kareemu	الواو waw	الأسماء الخمسة the Five Nouns	

Table. 3a. Variable nouns and their nominative end-case marks

	نوع الاسم Noun Type	الحالة الإعرابية وعلامة الإعراب Irab cases & End-Case Marks	
		النصب Accusative	Examples
معرب بالحركات diacritic end-case mark	الاسم المفرد بأنواعه singular جمع االتكسير broken plural جمع المؤنث السالم sound feminine plural	الفتحة fatha الفتحة fatha الفتحة fatha	ذاكر الولد الدرسَ thaakara al-waladu ad-darsa نصح الرجل الأولادَ nasaha ar-rajulu al-awlaada أسمع المحاضر الحاضراتِ asma'a al-muhadiru al-haderaati
معرب بالحروف letter end-case mark	جمع المذكر السالم sound masculine plural المثنى dual الأسماء الخمسة the Five Nouns	الياء ya الياء ya الألف alef	أكرمت المدرسة المعلمين akramati al-madrasatu al-mua'alimeena نصح المعلم الولدين nasaha al-mua'allimu al-waladayni رأى الرجل أبا الولد ra'a ar-rajulu aba al-waladi

Table. 3b. Variable nouns and their accusative end-case marks

	نوع الاسم Noun Type	الحالة الإعرابية وعلامة الإعراب Irab cases & End-Case Marks	
		الجر Genitive	Examples
معرب بالحركات diacritic end-case mark	الاسم المفرد بأنواعه singular جمع االتكسير broken plural جمع المؤنث السالم sound feminine plural	الكسرة ، الفتحة kasra, fatha الكسرة kasra الكسرة kasra	الولد في المدرسة al-waladu fi al-madrasati يلعبون في الملاعب yala'aboon fi al-malaa'ibi جاء الناس في الحافلات ja'a an-naasu fi al-hafilaati
معرب بالحروف letter end-case mark	جمع المذكر السالم sound masculine plural المثنى dual الأسماء الخمسة the Five Nouns	الياء ya الياء ya الياء ya	الكرة بين اللاعبين al-kuratu bayna al-la'ibeena أشار بيديه ashaara biyadayhi ذهب إلى أبي الولد thahaba ila abi al-waladi

Table. 3c. Variable nouns and their genitive end-case marks

type, and terminal letter. Tables 4a, 4b, and 4c show the possible irab cases and the corresponding end-case marks with examples. The reader will find it useful to compare the examples in Tables 4a, 4b, and 4c to observe the effect of the subjunctive and jussive verbal particles.

An invariable verb is a verb that does not experience any change in its end-case mark even if it is preceded by any of the verbal particles that affect the end-case of a verb. These verbs include all the past and imperative tense verbs, and the present tense verbs that are suffixed by the feminine "noon" or the energetic "noon" letter. The end-case mark for this type of verbs (called "al-bena'a" mark) differs for the same verb according to the suffix attached.

Particles are divided into two main classes: active and inactive [2 - 4, 6, 7]. The active particles are those that affect the case of succeeding words (nouns and verbs). The effect depends on the particle's type and the type of the affected word as well. The inactive particles have no effect on the succeeding word. The active particles are divided into three main classes: nominal, verbal, and nominal-verbal. The nominal articles put the succeeding nouns in the nominative, accusative, or genitive case according to the type of the particle and the end-case position of the noun. The verbal particles only precede a present tense verb and affect its case. The active verbal particles are divided into two main types: particles that put present tense verbs in the subjuntive case, and those that put present tense verbs in the jussive case. The nominal-verbal particles precede verbs and nouns, and include the conjunction ("al-atf") particles. These particles assign the end-case of the preceding word to the succeeding word. Fig. 2 shows the classification of particles and the types of particles under each class. Fig. 3 shows sample particles of different types. From the end-case point of view, all particles in Arabic are of the invariable type. That is, the particle terminal diacritic does not experience any change wherever it occurs in a sentence.

Sentences, from the end-case point of view, are divided into two classes: sentences with end-case position, and sentences with no end-case position. Tables 5a, and 5b show these two classes, sentence types of each class, and an example of each type.

Problem Definition

The problem is to design and implement an End-Case Analyzer (ECA) of Arabic sentences. The analyzer is to be integrated into a Natural Arabic Understanding System

الحالة الإعرابية وعلامة الإعراب Irab cases & End-Case Marks		نوع الفعل Verb Type	
Examples	الرفع Indicative		
يلعبُ الولد بالكرة yala'abu al-waladu bi-lkurati	الضمة damma	الفعل المضارع صحيح الآخر third-letter-sound present tense verb	معرب بالحركات diacritic end-case mark
يرى النجوم في النهار yara an-nujooma fi an-nahaari	الضمة المقدرة implicit damma	الفعل المضارع معتل الآخر بالألف third-letter-alef present tense verb	معرب بالحروف letter end-case mark
يدعو الناس إلى الخير yadoo an-nasa ila al-khayr	الضمة المقدرة implicit damma	الفعل المضارع معتل الآخر بالواو third-letter-waw present tense verb	
يقضي الطالب وقته في الدراسة yaqdi at-talibu waqtahu fi al-madrasati	الضمة المقدرة implicit damma	الفعل المضارع معتل الآخر بالياء third-letter-ya present tense verb	
الرجال يعملون في المصانع ar-rijaalu ya'amaloona fi al-masaani'i	ثبوت النون retention of noon	الأفعال الخمسة the five verbs	

Table.4a. Variable verbs and their indicative end-case marks

الحالة الإعرابية وعلامة الإعراب Irab cases & End-Case Marks		نوع الفعل Verb Type	
Examples	النصب Subjunctive		معرب بالحركات diacritic end-case mark
أراد أن يخرجَ araada an yakhruja	الفتحة fatha	الفعل المضارع صحيح الآخر third-letter-sound present tense verb	
يستطيع أن يرى النجوم yastateeo'u an yaraa an-nujooma	الفتحة المقدرة implicit fatha	الفعل المضارع معتل الآخر بالألف third-letter-alef present tense verb	معرب بالحروف letter end-case mark
خرج ليدعوَ الناس kharaja li-yado'ua an-nasa	الفتحة fatha	الفعل المضارع معتل الآخر بالواو third-letter-waw present tense verb	
أراد أن يقضيَ بالحق araada an yagdiya bil-haggi	الفتحة fatha	الفعل المضارع معتل الآخر بالياء third-letter-ya present tense verb	
عليهم أن يعملوا alayhim an ya'amaloo	حذف النون deletion of noon	الأفعال الخمسة the five verbs	

Table. 4b. Variable verbs and their subjunctive end-case marks

	نوع الفعل Verb Type	الحالة الإعرابية وعلامة الإعراب Irab cases & End-Case Marks	
		الجزم Jussive	Examples
معرب بالحركات diacritic end-case mark	الفعل المضارع صحيح الآخر third-letter-sound present tense verb	السكون skoon	لم يخرجْ lm yakhruj
معرب بالحروف letter end-case mark	الفعل المضارع معتل الآخر بالألف third-letter-alef present tense verb	حذف حرف العلة deletion of weak letter	لم يرَ النجوم lm yar an-nujooma
	الفعل المضارع معتل الآخر بالواو third-letter-waw present tense verb		لم يدعُ الناس lm yado an-nasa
	الفعل المضارع معتل الآخر بالياء third-letter-ya present tense verb		لم يقضِ بالحق lm yaqde bil-haggi
	الأفعال الخمسة the five verbs	حذف النون deletion of noon	لم يعملوا lm ya'amalbo

Table. 4c. Variable verbs and their jussive end-case marks

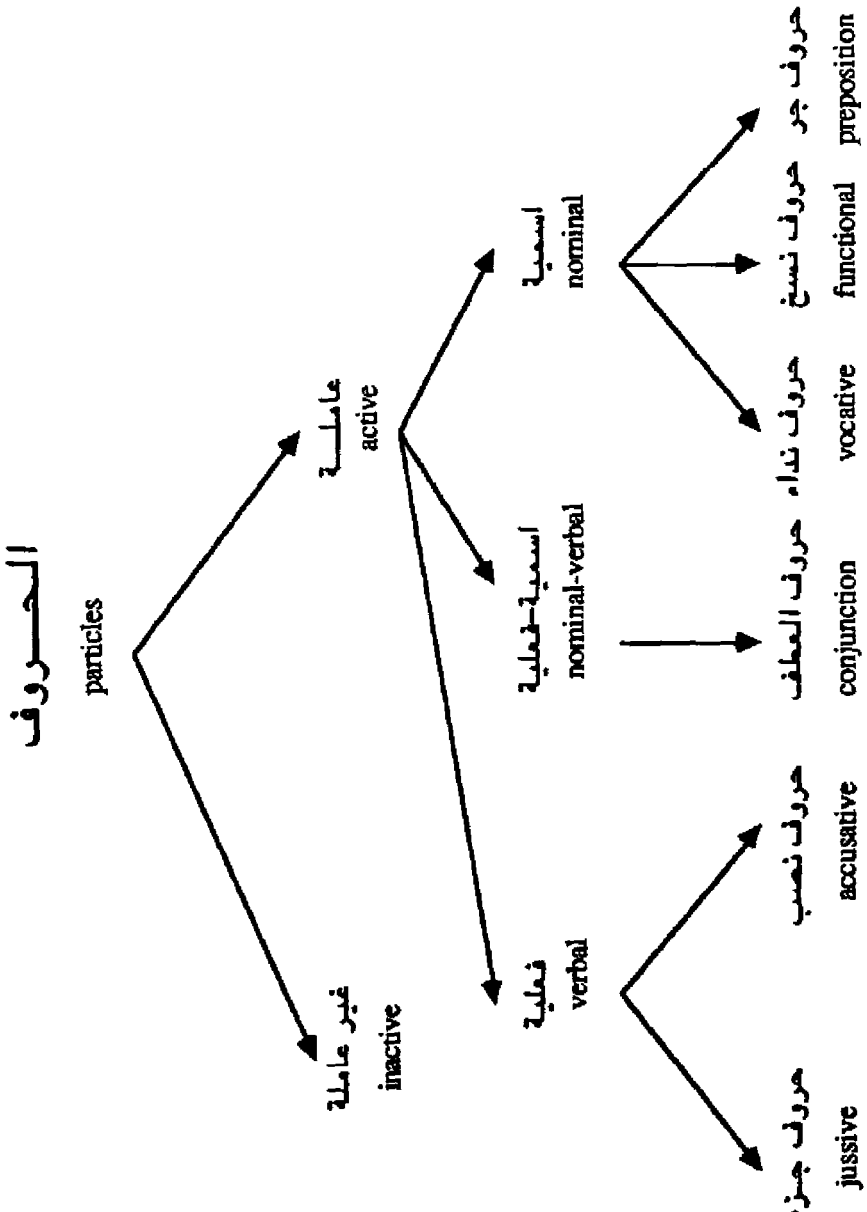

Fig. 2. Classification of arabic particles

inactive	active		
	verbal	nominal-verbal	nominal
lau لو	kay كي	wa و	inna إن
itha إذا	lm لم	thomma ثم	fi في
gad قد	hatta حتى	aw أو	ya يا

Fig. 3. Sample particles

Examples	الجمل التي لها محل من الإعراب Sentences with an End-Case Position
الولد يلعب بالكرة al-waladu yala'abu bil-kurati	الجملة الواقعة خبرا predicate sentence
ظننت الولد يضحك thanantu al-walada yadhaku	الجملة الواقعة مفعولا به object sentence
تظهر النجوم عندما تغيب الشمس tathharu an-nujoomu indma tagheebu a-shamsu	الجملة الواقعة مضافا إليه second part of the possessive construct sentence
من يجتهد فقد ينجح mn yajtahed fagd yanjahu	الجملة الواقعة جوابا لشرط جازم end-sentence of a conditional sentence with a jussive particle
خرج الرجل يدعو kharaja ar-rajulu yado'u	الجملة الواقعة حالا state sentence
جاء رجل يحب الناس ja'a rajulun yahebu an-nasa	الجملة الواقعة صفة modifier sentence
الطالب يذاكر درسه و ينجح at-talibu yuthkiru darsahu wa yanjahu	الجملة التابعة لجملة لها محل من الإعراب dependent sentence of a sentence with an end-case position

Table. 5a. End-case classification of arabic sentences with end-case position

Examples	الجمل التي لا محل لها من الإعراب Sentences with no End-Case Position
الولد يلعب بالكرة al-waladu yala'abu bil-kurati	الجملة الابتدائية predicate sentence
إنه لو تعلمون أمر عظيم innahu law ta'alamoona amrun atheemun	الجملة المعترضة object sentence
أوحينا إليه أن اصنع الفلك awhaina ilaihi an esna'ai el-fulka	الجملة التفسيرية second part of the possessive construct sentence
إذا ذاكرت نجحت itha thaakarta najahta	الجملة الواقعة جوابا لشرط غير جازم end-sentence of a conditional sentence with a jussive particle
والله إن محمد لصادق wa-llahi inna mohammadun lasadigun	الجملة الواقعة جوابا لقسم state sentence
جاء الذي يقول الحق ja'a alathi yagool al-hagga	الجملة الواقعة صلة لاسم موصول modifier sentence
الولد يلعب و البنت تدرس al-waladu yala'abu wa al-bintu tadrusu	الجملة التابعة لجملة لا محل لها من الإعراب dependent sentence of a sentence with an end-case position

Table. 5b. End-case classification of arabic sentences with no end-case position

(NAUS) that is under development. Given the output of the morphological and the syntactic analyzers of NAUS, ECA is to produce the end-case analysis of the sentence under consideration. The morphological analyzer determines the characteristics of the words of the sentence; the syntax analyzer detrmines the syntactic functions (e.g., subject of verb, verb, object of verb, ... etc) of the constructs of the sentence.

The purpose of ECA is to determine the end-case characteristics of all the constructs in the sentence being processed. In addition, ECA is to determine the irab of the sentence as a whole, and the irab of the constituent subsentences, if any. The end-case characteristics [2 - 4] include the following:

a) The end-case position, which represents the syntactic function of the construct in the sentence.
b) The explicit/implicit irab case of each construct of the sentence.
c) The end-case mark for each construct.

Literature Survey

With the recent trend of computer Arabization, research has been directed towards analysis of Arabic as a natural language. However, most of this research mainly concentrated on the fields of morphological analysis [8 - 13] and syntactic analysis [13, 15, 16].

The research work in the end-case analysis field is very little. In [17], a system capable of performing lexical and end-case analysis of constructs is described. This system is a lexical analyzer with the first elementary stage of an end-case analyzer. The system can only analyze isolated single constructs; it does not determine the word's end-case position in a sentence. In another effort [18], a system was based on supplying the system with the grammatical rule in some special form before asking the system for an analysis that depends on that rule. The system is only capable of producing the end-case positions of words of simple non-imperative sentences. It does not produce end-case analysis of whole sentences. A third effort [19] presented an end-case analyzer of simple sentences only. The system was based on determining the compatibility of pairs of consecutive words. Although the system determines the end-case position and the end-case mark of the words in the given sentence, it is restricted to the analysis of simple sentences only, and does not address end-case analysis of sentences as whole units.

Design of ECA

NAUS accepts an Arabic sentence from the user, passes it through the syntactic and morphological analyzers and hence forwards it along with the syntactic and morphological output to ECA. The input to ECA consists of:

a) The sentence as a sequence of words.
b) the morphological characteristics, which specify the type of word (noun, verb or particle), person, number, gender, tense (past, present, or imperative for verbs), type and identity of affixes (prefix, infix, or suffix), root and derivation pattern (for nouns and verbs).
c) The syntactic structure of the sentence. The syntactic structure includes the general syntactic functions of the constructs of the sentence, and the type of the sentence.

The output of ECA includes the following end-case characteristics for each construct of the given sentence:

1- The type (variable and invariable) of the construct,
2-. The end-case position of the construct,
3- The irab case of the construct,
4- The end-case mark of the construct, and
5- For sentences, the implicit end-case position of the whole sentence.

When developing an end-case analyzer as a component of a natural Arabic understanding system, the task can be divided into two main functions: invocation of the analysis within the system and performing the analysis, which is an independent activity given all required input. Thus, ECA is designed to consist of two main modules: the invoker and end-case generator.

The development of the invoker involves incorporation of invocations of the required rules of the end-case generator into NAUS. Since syntax and end-case in Arabic are highly interrelated, the structure of the syntax analyzer greatly influences that of the invoker. As the syntax analyzer recognizes sentences, the corresponding end-case rules can be invoked.

Furthermore, since the syntax analyzer of NAUS is a top-down parser, compound structures such as noun phrases are recognized as units and must be decomposed by the invoker into words in order to invoke the corresponding end-case rules of the end-case

generator. This approach is further imposed by the fact that the irab cases for some end-case positions, e.g., subject, predicate, verb, and sentence are context-sensitive (due to the presence or absence of active particles and imperfect verbs). In this case, the end-case analysis invocation must be placed in the module that recognizes the context in which the construct occurs. For the context-insensitive cases, e.g., subject of verb and object of verb ("fa'el" and maf'ool"), The corresponding invocations for the end-case rules of the end-case generator can be placed in the rule (in the syntax analyzer) that recognizes the syntactic functions of the construct, thus minimizing the number of occurrences of such invocations.

The end-case generator consists of four main modules for nouns, verbs, particles and sentences respectively. Based on the irab cases of the noun, the noun module consists of three submodules (nominative, accusative and genitive) which determine the irab case and end-case mark of single words in accordance with Tables 1a, 1b, and 1c. The verb module consists of three submodules: imperative, past tense, and present tense. The imperative verbs are of the invariable type, but are of two kinds: perfect and imperfect. Since the end-case analysis of the verb itself does not depend on the fact that a verb is perfect or imperfect, both types of verbs are processed in the same way except for determining the verb's type. Accordingly, only one module, with a parameter for the verb's type, suffices to process all the imperative verbs. Past tense verbs are also of the invariable type and are further classified into two subclasses: active and passive voice. Accordingly, we have two modules for the past verbs. The present tense class is analogously classified into two subclasses: active, and passive voice. From the end-case point of view, however, the present tense class is subdivided into variable and invariable verbs. Varable verbs are further divided into three subdivisions: indicative, subjunctive, and jussive. Applying this division to the two classes of present tense, namely, active and passive voice classes, we have six modules for variable present tense verbs. For invariable present tense verbs, there are two modules (for active and passive voice). Fig.4 shows the design structure of the verb module. Each of the eleven verb modules encodes the corresponding Arabic end-case grammar rules for the corresponding type of verb. The module is designed to process any kind of particle according to its identity. The input to the module is the identity of the particle and the particle itself. The output of the module includes: the type of particle, whether it is an active or a inactive particle, and its action in the sentence if it is an active particle. The rules for end-case analysis of sentences as whole units are encoded according to Table 5.

Fig. 5 shows the overall design of the system. Fig. 6 shows the end-case analyzer embedded into NAUS.

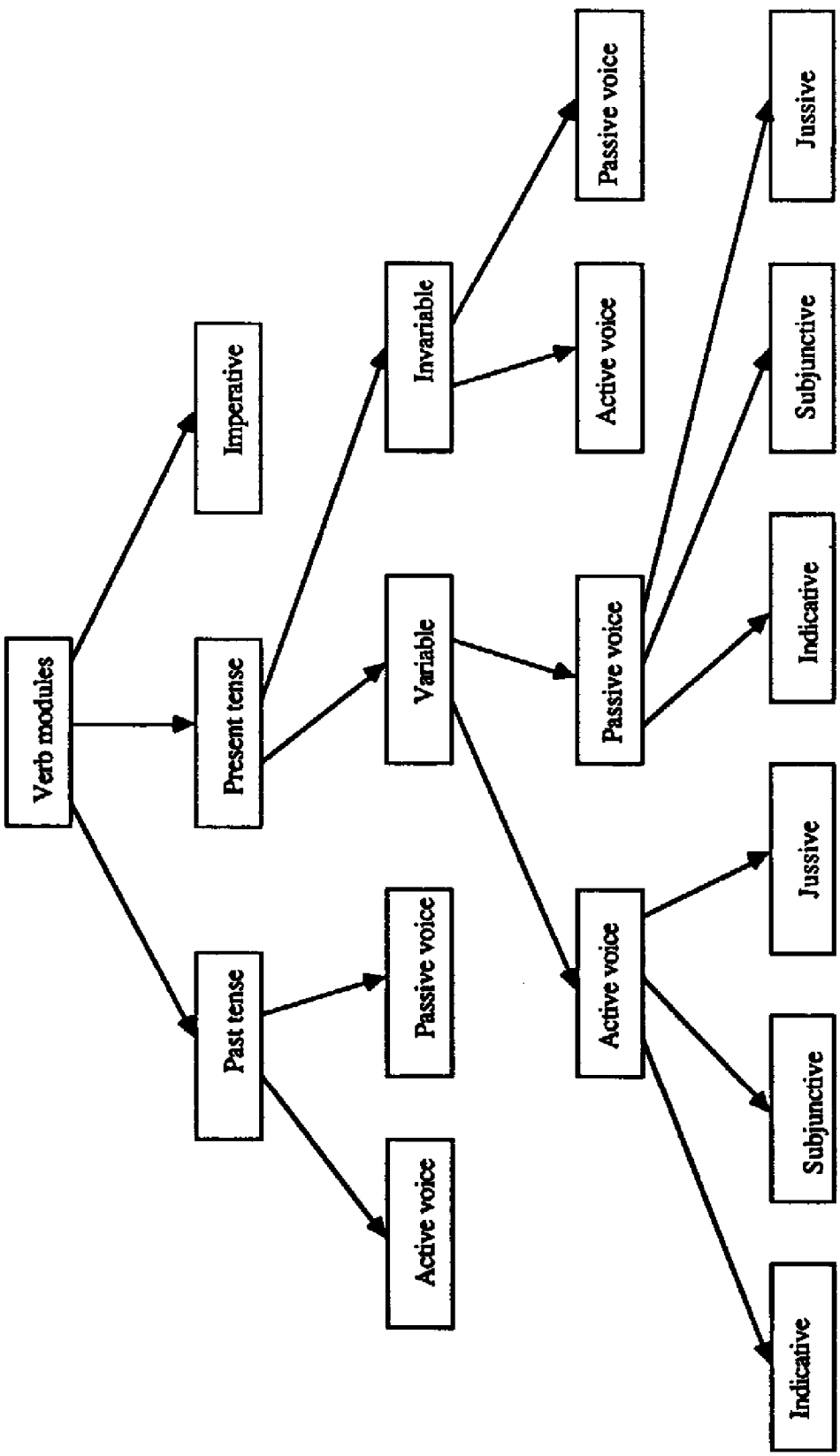

Fig. 4. Verb module design

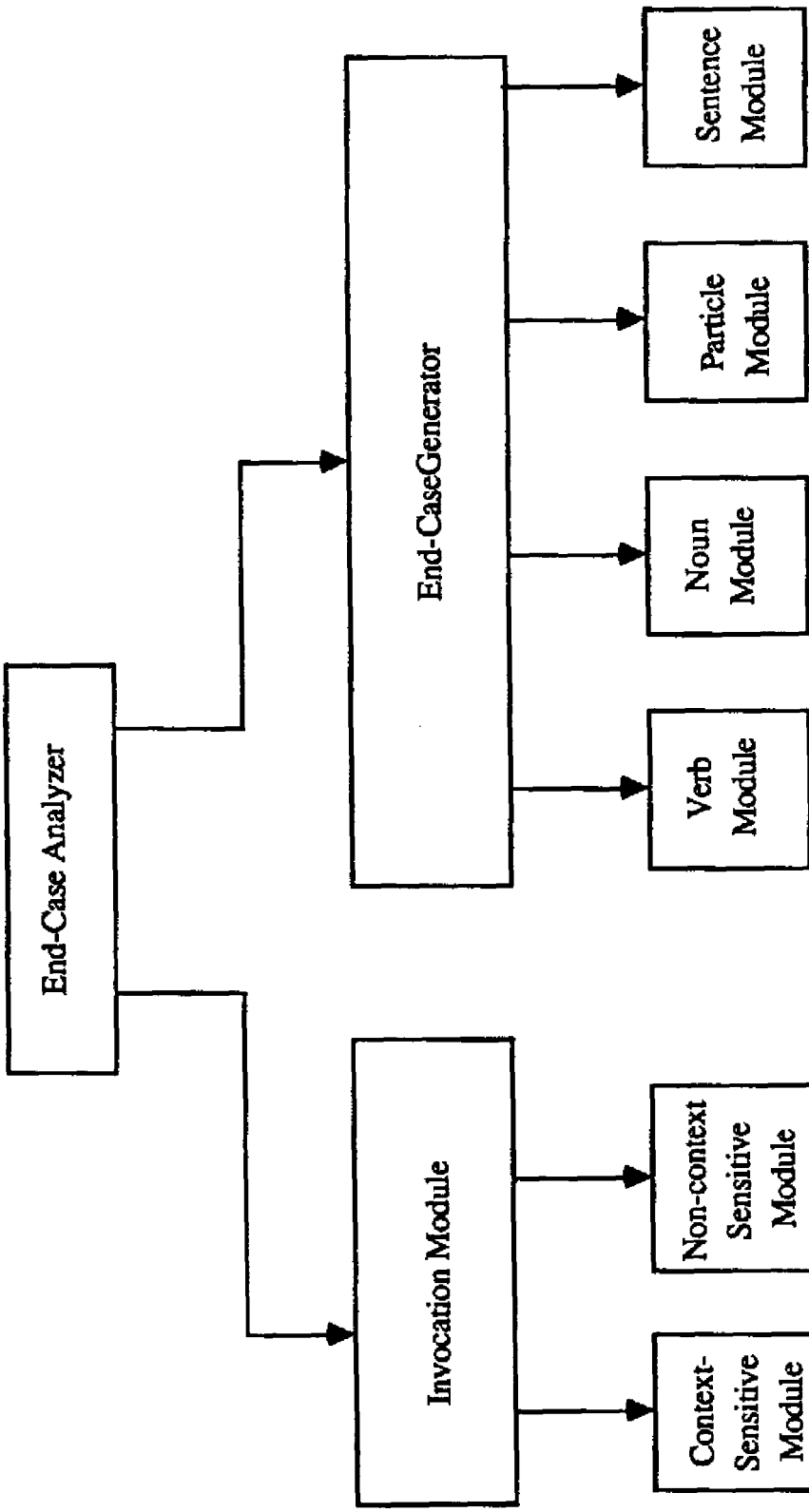

Fig. 5. Design structure of ECA

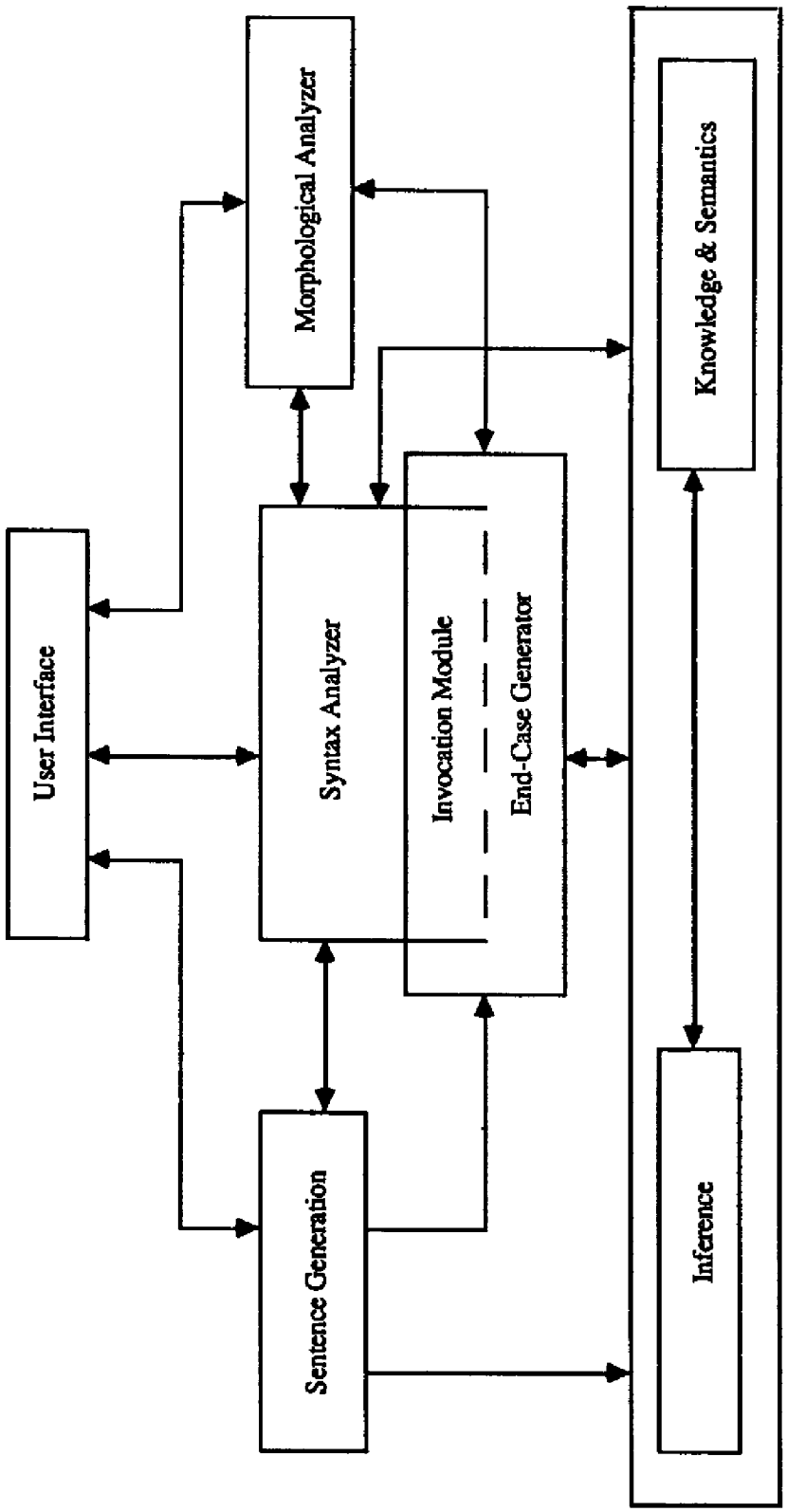

Fig. 6. ECA embedded into NAUS

Implementation of ECA

ECA was implemented in Prolog. The incoker was implemented by inserting the appropriate calls within the syntax analyzer. The number of rules used for decomposing compound structures is 51. The number of rules of the end-case generator is divided as follows: 132 for nouns, 52 for verbs and 55 for particles. Fig. 7 shows some examples of calling the end-case rules for context-sensitive and context-insensitive constructs.The predicate bfb_feil of the syntax analyzer recognizes a verbal sentence consisting of a verb, subject of verb, and object of verb which are recognized by the predicates feil, fael, and mfol respectively. As the end-cases of the subject of verb and object of verb are context-insensitive, their corresponding end-case calls are made in the predicates fael and mfol (not shown in the example). The predicate bfm represents a verbal sentence consisting of an imperfect verb, recognized by the predicate kan, and a nominal sentence (consisting of a subject and a predicate) recognized by the predicate bab. End-case analysis of the imperfect verb is performed in the context (not shown) in which the predicate bfm occurs since it is a context-sensitive verb. The end-case of the subject is nominative and is recognized by the predicate asma_marfo, and that of the predicate of the nominal sentence is accusative and is recognized by the predicate asma_mnsb. Both cases are context-sensitive and the calls are made in the context in which they occur. The predicate afm illustrates a case in which the end-case of a sentence represented by the predicate afb or bfb is context-sensitive and is determined by the predicate irab-jmlh as the predicate of an imperfect verb.

The syntax analyzer of NAUS divides the sentences of Arabic into eight classes that cover the whole spectrum of sentence structures. ECA was run for all classes of sentences. Appendix A shows a partial sample of the results. With respect to single words, every irab case was tested (in sentences) and the correct analysis verified. Fig. 8 shows an output sample of ECA with corresponding translation and transliteration. The first line is the sentence being analyzed. The sentence is recognized by the predicate afm shown in Fig. 7. The sentence consists of the imperfect verb "awshaka" (recognized by the predicate raja_awshk), the subject of "awshaka", (an implicit pronoun recognized by the predicate mbtda), the particle "an" (recognized by the predicate "an"), and the corresponding predicate (the verbal clause "yanzaliqa fi al-hofrati" recognized by one of the predicates afb or bfb). Note that the verb of the verbal clause is required (by predicate xtrct) to be in the present tense. As the verbal clause is parsed, end-cases of the particle "an", the prepositional phrase "fi al-hofrati" and the subject of verb, an implicit pronoun, are produced as shown in the second, third, fourth and fifth lines by the corresponding recognizing predicates, since they are context-insensitive. However, note

that the end-case of the present tense verb "yanzaliqa" is not performed within the

```
% an example of calls for context insensitive constructs

bfb_feil([L1:L2],[Nom1,Nom2,Nom3,Nom4],Rem) :-
                          [!feil(L1,[Typ,$مضارع$],Nom1,Lsfx1)!],
                          fael(Lsfx1,L2,Nom2,Jhmn2,L3,Lsfx2),
                          mfol(Lsfx2,L3,Nom3,Jhmn3,L4,[$$]).


% examples of calls for context sensitive constructs

bfm([L1:L2],[Nom1,Nom2],Rem) :-
                            kan(L1,Nom1,Jh1),
                            bab(L2,Nom2,Rem),
                            [_,[_,Y2],[_,Z2]]=Nom2,
                            asma_marfo($كان اسم$,Y2),
                            asma_mnsb($كان خبر$,Z2).

afm([L1:L2],[Nom1,Nom2,Nom3,Nom4],Rem) :- raja_awshk(L1,Nom1),
                        mbtda(L2,Nom2,Jhmn2,[L3:L4]),an(L3,Nom3),
                        xtrct(2,Nom4,[[_,[_,_,_,$معارع$|_]]|_]),
                        (afb(L4,Nom4,Rem);bfb(L4,Nom4,Rem)),
                        asma_marfo($الفعل الناقص اسم$,Nom2),
                        [X,[[F2,Y2]|_]]=Nom4,
                        f_mansob(F2,Y2,_),
                        irab_jmlh($ في محل نصب خبر الفعل الناقص $,X).
```

Fig. 7. Example calls of the invoker within the syntax analyzer

```
1   (أوشك أن ينزلق في الحفرة
2   (أن : حرف نصب ينصب الفعل المضارع)
3   (في : حرف جر)
4   (الحفرة : اسم مجرور وعلامة الجر الكسرة الظاهرة على آخره)
5   (هو : فاعل : ضمير مستتر في محل رفع)
6   (هو : اسم الفعل الناقص : ضمير مستتر في محل رفع)
7   (ينزلق:فعل مضارع منصوب وعلامة النصب الفتحة الظاهرة على آخره)
8   (فعلية أدواتية بسيطة : في محل نصب خبر الفعل الناقص)
9   (أوشك:فعل ناقص ماضي مبني على الفتح)
1   (أدواتية فعلية مركبة : جملة ابتدائية لا محل لها من الاعراب)


1   Translation: "He was about to fall in the ditch"
1   Transliteration " awshaka an  yanzaliqa fi al-hufrati".

2   (an : subjunctive verbal particle)
3   (fi : prepositional genitive particle)
4   (al-hufrati : genitive noun, end-case mark is kasra)
5   ( subject of verb (yanzaliqa) :
                  implicit pronoun in the nominative case)
6   ( subject of imperfect verb (awshaka) :
                    implicit pronoun in the nominative case)
7   (yanzaliqa : present tense verb in the subjunctive case,
          end-case mark is fatha)
8   (simple particle verbal sentence :
     predicate of imperfect verb, end-case : implicit fatha)
9   (awshaka : imperfect past tense verb, end-case mark is fatha)
1   (compound particle verbal sentence :
              independent sentence with no end-case position)
```

Fig. 8. A sample output

predicate afb or bfb because it is context-sensitivej. Next, the predicate asma_marfo proxuces the end-case of the subject of "awshaka" (line 6). The predicate f_mansob produces the end-case of the verb "yanzaliqa" in line 7. Note that the case of the verb is subjunctive because it is preceded by the subjunctive particle "an". Next, the predicate irab_jmlh produces the end-case of the predicate clause of "awshaka" (line 8). The last two lines (produced in the predicate that invoked the predicate bfm) in Fig. 8 show the end-case analysis of the verb "awshaka" and that the whole sentence has no end-case because it is independent.

The source code size of the end-case analyzer, including documentation, was about 130 KB. However, memory usage by the analyzer, including the calls inserted within the syntactic analyzer, was about 84 KB only. The end-case generator consists of 290 rules. In order to measure the performance of ECA and avoid the effects of other modules of NAUS, a driver was used to run ECA separately and simulated morphological and syntactic input was used. Although the analyzer was run in interpreter mode, the response time was a few seconds. Given the correct morphological and syntactic input, ECA always produces the correct end-case analysis, otherwise ECA returns failure.

Conclusion

In this research we have designed and implemented an end-case analyzer (ECA) of Arabic sentences. The end-case analyzer encodes the Arabic grammar rules of irab and the effects of applying these rules to the constructs of sentences. The analyzer has been built to be used as a component of a Natural Arabic Understanding System (NAUS). ECA uses the output of the morphological and syntactic analysis modules of NAUS to determine the end-case position and the end-case mark for each construct in the sentence. ECA consists of two main components, the end-case generator and the invoker. The end-case generator encodes the rules of irab for all constructs of the Arabic sentence. The invoker initiates the relevant end-case rule from within the syntax analyzer of NAUS.

ECA interacts with the morphological and syntax analyzers to resolve ambiguities and determine the correct analysis of sentence. ECA rejects the morphological or syntactic input if no end-case rule applies. In this case, the morphological and syntax analyzers are recalled (through backtracking) to produce another possible analysis. The process is repeated until an acceptable analysis is produced or the sentence is rejected. The extent to which ambguity can be resolved is not comprehensive since ambguity can be semantic which is another research problem.

ECA can be easily adapted towards any other related system or application by only restructuring the invoker to fit within the application under consideration. As an immediate application of ECA, it has been slightly modified to serve as an end-case synthesizer (in addition to analysis) in NAUS for sentence generation. For future work, the analyzer can be integrated into some other Arabic applications such as text generation, speech analysis/synthesis, systems for teaching Arabic, translation systems, and systems for checking and correcting grammatical errors in written and spoken Arabic.

Appendix A

```
الرجل صاحب المنزل كريم
(الرجل : مبتدأ : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره)
(صاحب : صفة وهو مضاف : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره)
(المنزل : مضاف اليه : اسم مجرور وعلامة الجر الكسرة الظاهرة على آخره)
(كريم : خبر : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره)
(اسمية بسيطة : جملة ابتدائية لا محل لها من الاعراب)

ذهب الرجل يدعو
(هو : فاعل : ضمير مستتر في محل رفع )
(الرجل : فاعل : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره)
(يدعو :يدعو :فعل مضارع مرفوع وعلامة الرفع الضمة المقدرة للثقل )
(فعلية بسيطة : الجملة في محل نصب حال)
(ذهب :ذهب :فعل ماضي مبني على الفتح )
(فعلية مركبة 1 : جملة ابتدائية لا محل لها من الاعراب)

كان الولد صالحا
(الولد : اسم كان : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره)
(صالحا : خبر كان : اسم منصوب وعلامة النصب الفتحة الظاهرة على آخره)
(كان :كان :فعل ناقص ماضي مبني على الفتح )
(فعلية مركبة : جملة ابتدائية لا محل لها من الاعراب)

الرجل والضيف في المنزل
(في : حرف جر )
(المنزل : اسم مجرور : اسم مجرور وعلامة الجر الكسرة الظاهرة على آخره)
(الرجل : مبتدأ : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره)
(و : حرف عطف )
(الضيف : معطوف : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره)
( جار ومجرور : شبه جملة في محل رفع خبر)
(واو الحالية اسمية بسيطة : جملة ابتدائية لا محل لها من الاعراب)


ان الرجل كريم
(ان : حرف ناسخ مشبه بالفعل ناصب للاسم رافع للخبر )
(الرجل : اسم ان : اسم منصوب وعلامة النصب الفتحة الظاهرة على آخره)
(كريم : خبر ان : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره)
(واو الحالية اسمية بسيطة : جملة ابتدائية لا محل لها من الاعراب)

الرجل الذي رايته محمد
(ت : فاعل : ضمير متصل في محل رفع)
(ه : مفعول به : ضمير متصل في محل نصب )
(رايته :راي :فعل ماضي مبني على السكون لاتصاله بضمير رفع متحرك )
(فعلية بسيطة : جملة صلة لا محل لها من الاعراب )
(الرجل : مبتدأ : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره)
(الذي : صفة : اسم موصول مبني على السكون في محل رفع)
(محمد : خبر : اسم علم مرفوع وعلامة الرفع الضمة الظاهرة على آخره)
(اسمية مركبة : جملة ابتدائية لا محل لها من الاعراب)
```

البيت نظيف والحديقة جميلة
{البيت : مبتدأ : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره}
{نظيف : خبر : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره}
{والحديقة : مبتدأ : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره}
{جميلة : خبر : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره}
{اسمية بسيطة : جملة ابتدائية لا محل لها من الاعراب}
{اسمية بسيطة : الواو حرف عطف والجمله
معطوفة على جمله ابتدائية لا محل لها من الاعراب}

نعم الصديق الذي يجود في الضيف
{في : حرف جر }
{الضيف : اسم مجرور : اسم مجرور وعلامة الجر الكسرة الظاهرة على آخره}
{و : فاعل : ضمير مستتر في محل رفع }
{يجود :يجود :فعل مضارع مرفوع وعلامة الرفع الضمة الظاهرة على آخره}
{فعلية ادوائية بسيطة : جملة صلة لا محل لها من الاعراب }
{نعم :نعم: فعل ماضي مبني على الفتح }
{الصديق : فاعل : اسم مرفوع وعلامة الرفع الضمة الظاهرة على آخره}
{ الجملة من الفعل والفاعل : في محل رفع خبر مقدم }
{الذي : مبتدأ موخر : اسم موصول مبني على السكون في محل رفع}
{ادوائية فعلية مركبة : جملة ابتدائية لا محل لها من الاعراب}

ان تذاكر تنجح
{انت : فاعل : ضمير مستتر في محل رفع }
{تذاكر :تذاكر : فعل الشرط ، فعل مضارع مجزوم وعلامة الجزم السكون }
{انت : فاعل : ضمير مستتر في محل رفع }
{تنجح :تنجح : جواب الشرط ، فعل مضارع مجزوم وعلامة الجزم السكون }
{فعلية بسيطة : جملة جواب الشرط لا محل لها من الاعراب }
{ان : حرف شرط جازم يجزم فعلين }
{ادوائية فعلية مركبة : جملة ابتدائية لا محل لها من الاعراب}

References

[1] Abdo, D. A. *"A Course in Modern Standard Arabic."* Beirut:: Almaktaba Alhadeetha, 1962.

[2] Ghalayeeni, M. *Jame'a Al-Dorus Al-Arabia.* (in Arabic), 22nd Edition, Beirut : Almaktaba Alaasriyah, 1979/1409 H.

[3] Abo-Mesleh, K. *Al-Kamel fi Al-Nahow wa Al-Sarf.* (in Arabic, Four Volumes). 2nd edition , Beirut: The Modern Bookshop, 1984.

[4] Dahdah, A. *Mo'ajam Kawa'ed Al-Lugha Al-Arabia.* (in Arabic), 2nd Edition, Beirut::Maktabat Lebanon , 1985 .

[5] Cantarino, V. *Syntax of Modern Arabic Prose - The Simple Sentence.* volume 1. Bloomington: Indiana University Press, 1974.

[6] Al-Tunaji, M. *Mo'jam Al-Adawat Al-Nahaweiah.* (in Arabic), 4th edition, Damascus: Dar Al-Feker, 1968.

[7] Al-Ansary, Ibn Hesham. *Al-Irab an Kawa'ed Al-Irab.* (in Arabic, verified by A.F. Nabil) 1st Ed., Riyadh:Riyadh University (KSU), 1981/1401 H.

[8] Helal, Y. "A Morphological Analysis for Arabic Speech." *A Workshop about Computerized Arabic Processing*, Kuwait: (April 1985).

[9] Gheith, M. and El-Sadany, T. "An Arabic Morphological Analyzer on a Personal Computer." *Proceeding of The First K.S.U. Symposium on Computer Arabization*, (April 1987).

[10] Al-Bawab, M.; Meer-alam, y. and Al-Tayyan, M. " A Computrized Morphological Arabic System." *The Regional Conference on Informatics and Arabization*, Tunis, (1988).

[11] Al-Authman, A. "A Morphological Analyzer for Arabci." *M.S. Thesis*, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia, (1989).

[12] Al-Zarroq, A." A Morphological Anlyzer for Derived Arabic Words." *The Second Conference on Arabic Computational Linguistics*, Kuwait, (Nov. 1989).

[13] Beesley, K."Computer Analysis of Arabic Morphology: a Two-Level approach with Detours. " *Proceedings of The Third Annual Symposium on Arabic Linguistics*, University of Utah, (1989).

[14] Al-Saied A. N. "The Computerized Analysis of Arabic Language Systems- Part One: The Formation of Arabic Grammar as a Context Free Grammar." *The 1st Conference for Computers*, Kuwait, (March 1989).

[15] Gheith, M. and Aboul Ela, M."A Computer Based Arabic Syntax Analyzer." *The 11th National Conference*, Dhahran, Saudi Arabia, (March 1989).

[16] Khayat, M. G. And Al-Jabri, S. K."A Model for a Syntax Analyzer of Arabic." *The 12th National Conference*, Riyadh, Saudi Arabia, (October 1990).

[17] Al-Bawab, M. and Merayati, M. "A Computerized Morpho-Syntax System for Arabic." *The 2nd Conference on Arabic Computational Linguistics*, Kuwait, (Nov. 1989).

[18] Al-Saied, A. N. and Al-Garah, S. "The Computerized Analysis of Arabic Language Systems - Part Two: An Elementary Model for a Computerized End-Case Analyzer for Non-Imperative Nominal Arabic Senstences." *The 2nd Conference on Arabic Computational Linguistics*, Kuwait, (Nov. 1989).

[19] Al-Sheshaini, H. and Al-Najjar, A. " Two Computerized End-Case Analyzers for Arabic Sentences." *The 2nd Conferenc on Arabic Computational Linguistics*, Kuwait, (Nov. 1989).

محلل إعرابي للجملة العربية

ا.د. السوادي و محمد غزالي خياط *

*جامعة الملك فهد للبترول والمعادن، الظهران و * قسم الإلكترونيات وهندسة الحاسب، جامعة الملك عبدالعزيز، جدة، المملكة العربية السعودية*

(قدم للنشر في ٦/ ٩/ ١٩٩٣م، وقبل للنشر في ٢١/ ٩/ ١٩٩٤م)

ملخص البحث. إن فهم اللغات الطبيعية من أهم مجالات البحث في علم الحاسب الآلي. وبالرغم من أن الصرف والتركيب لهما أهمية في اللغات الطبيعية، إلا أن الإعراب، وخاصة في اللغة العربية، لهو عنصر ذو أهمية بالغة في تحديد دلالة الجملة.

ويستعرض هذا المقال تصميم وتنفيذ محلل إعرابي وذلك في إطار نظام لفهم اللغة العربية يسمى "نوس". ويتكون المحلل إعرابي من عنصرين رئيسين هما: المنتج والناشد. ويقوم المنتج بتحديد إعراب الجملة ومكوناتها وفقاً لقواعد النحو التي تم تمثيلها كقواعد لغة برولوج الشرطية. وقد تم تطوير الناشد بشكل استدعاءات للمنتج في المحلل التركيبي لنظام "نوس". ويتم حاليًا استخدام المحلل الإعرابي على جهاز حاسب إلى شخصي مزود بنظام تعريب.

إن تصميم المحلل الإعرابي كوحدات يتيح استخدامه في نظام "نوس" بالإضافة إلي التطبيقات الأخرى مثل إنشاء الجمل وتحليل وتركيب الكلام ونظم الترجمة وتصحيح الأخطاء اللغوية.